7N-64-CR
210786
41P.

ICASE REPORT

INITIALIZATION BY COMPATIBLE BALANCING

Michael Ghil

Report Number 75-16
August 22, 1975

INSTITUTE FOR COMPUTER APPLICATIONS

IN SCIENCE AND ENGINEERING

Operated by the

UNIVERSITIES SPACE RESEARCH ASSOCIATION

at

NASA'S LANGLEY RESEARCH CENTER

Hampton, Virginia

INITIALIZATION BY COMPATIBLE BALANCING

Michael Ghil

Abstract

The initialization problem is defined as the problem of obtaining the initial data that are required in order to solve a well-posed initial-and-boundary value problem for the equations of large-scale dynamical meteorology. In our treatment of the problem we assume that complete information is available at a given instant on the atmospheric temperature and surface-pressure fields, as well as on their time derivatives; we study a procedure for computing the horizontal velocity field at the instant of interest based on the assumed information.

This procedure relies on the exact mathematical treatment of the differential equations of atmospheric motion, rather than on perturbation-type or numerical methods; therefore the initial state given by the known temperature field and the derived velocity field should be fully compatible with the equations describing the time evolution of the system and not give rise to spurious noise components of the motion.

In the case of an atmospheric model governed by the linearized shallow-fluid equations, the diagnostic equations obtained by our procedure for the wind field are of uniformly elliptic type and an error analysis relating the accuracy of the results to that of the data is possible. In the case of a model described by the full non-linear shallow-fluid equations, the diagnostic equations we obtain for the wind field are of mixed type and related to the

classical balance equation; the ellipticity condition we derive for the former is similar to the well-known ellipticity condition for the latter. The important feature of this condition is that it does not depend on the partial derivatives of the horizontal velocity components. Finally, in the case of a primitive-equation model our diagnostic equations for the horizontal velocity field are similar to those for the non-linear shallow-fluid model, though somewhat more complicated. Their type is discussed, and a general condition to determine ellipticity is given. This condition does not appear to be so easily reduced to a simple form, independent of velocity derivatives, as in the previous case.

ACKNOWLEDGEMENT

The work reported on herein was started at the Goddard Institute for Space Studies during the summer of 1972. The author greatly benefited from conversations he had at different times with Prof. J. G. Charney, with Drs. S. H. Schneider, T. Gal-Chen, M. Halem, A. Kasahara, W. Washington, R. Bleck, K. Miyakoda, H. C. Davies and with Mr. R. E. Turner. Discussions with Professors P. D. Lax and E. Isaacson were particularly fruitful.

This research was supported in part while the author was in residence at the Courant Institute for Mathematical Sciences, New York University, by the Advanced Research Projects Agency under Grants No. DA-ARO-D-31-124-72-G113 and No. AFOSR-74-2728, and by the Goddard Institute for Space Studies under Grant No. NASA-NSG-5034; it was also supported in part under NASA Grant NGR 47-102-001 while the author was in residence at ICASE, NASA Langley Research Center, Hampton, VA 23665.

1. Introduction

One of the main difficulties in improving short-term weather forecasting based on numerical integration of the primitive equations[1] is the lack of initial data. The primitive equations in cartesian coordinates, say, form a hyperbolic system of first-order partial differential equations (PDEs), and weather prediction involves solving numerically an initial-and-boundary value problem for these equations. We shall not discuss here the boundary conditions to be imposed at the upper and at the lower boundary of the atmosphere. The required initial data at a given time $t = t_0$, to be specified throughout the global atmosphere, are the horizontal velocity components u, v, the density ρ and the temperature T. The complete and sufficiently accurate specification of such initial data constitutes the initialization problem.

Thus the initialization problem naturally falls into two parts: (i) the completeness, and (ii) the accuracy of the initial data. In discussing the different aspects of this problem, we shall try to distinguish between (a) the physical behavior of the atmosphere, (b) the mathematical (differential) models describing this behavior, and (c) the numerical (difference) models used to approximate a given mathematical model.

The question of accuracy of the initial data is particularly critical because of the extreme sensitivity of presently available numerical models to changes in initial data. In other words, small relative errors in the initial data result within a few hours of simulated time in large relative errors in the solution (Charney *et al.*, 1966). These errors manifest themselves mainly in a form identified as the numerical counterpart of fast inertia-gravity waves (Hinkelmann, 1951, Charney, 1955).

[1]For simplicity we shall call primitive equations the Euler equations of fluid dynamics, with the vertical momentum equation replaced by the hydrostatic assumption, and with the energy equation in its isentropic form.

In the atmosphere inertia-gravity waves (IGWs) are present only with small amplitudes, because of the mechanism of geostrophic adjustment; this adjustment results in a so-called balanced state, in which slow, meteorologically significant, motions are predominant. Therefore, it was thought that a balanced numerical solution with smaller IGW-like errors could be produced by requiring the initial data to satisfy a time-independent compatibility condition approximately expressing the property of balance. Thus the idea of a time evolution free of large inertia-gravity waves was tied to that of an instantaneous relation, containing no time-derivatives, between wind field and mass field; this relation was called the balance equation.

A number of forms of this balance equation (Ellsaesser, 1968, Haltiner, 1971, pp. 60-61) have been formulated and used in the initialization of primitive-equation models. Their use did not prevent, however, the appearance and exaggerated growth of IGW-like components of the solution in the numerical integration of these models (Nitta and Hovermale, 1969, Morel et al., 1971). The mathematical reason for this is the fact that all these forms of the balance equation involve in their derivation from the equations of motion the neglecting of certain terms; therefore such a balance equation is consistent with the equations of motion only in an approximate, asymptotic sense, rather than in an exact one.

As synoptic balancing did not seem to be successful with primitive-equation models, a form of dynamical balancing was proposed by Nitta and Hovermale (1969); the basic idea was to reduce the amplitude of spurious IGW-like phenomena in the model by using the dispersive and dissipative properties of the model itself, i.e., let the model simulate the geostrophic adjustment process in the atmosphere. A related idea was put forward by Charney et al. (1969) for using the time-dependent model to solve the completeness-of-data question.

Indeed, no observational system in use at present can supply at a given time (synoptically) all of u, v, ρ, T over a uniform grid, like the ones used in general circulation models (GCMs). However, it seems reasonable to expect through the use of satellite technology better data coverage for temperature than for winds. The suggestion then, following some ideas already found in Smagorinsky and Miyakoda (1969), was to use the history of the temperature field T and that of the surface-pressure field p_s from $t=t_0$ to $t=t_1$ in order to infer the horizontal wind field (u, v) at $t=t_1$. In fact, in a primitive-equation model T and p_s completely determine all thermodynamic variables and thus the mass field, since the equation of state and the hydrostatic equation immediately yield p and ρ everywhere.

The concrete procedure proposed by Charney et al. (1969) was to start the numerical integration of the equations with the "correct" temperature data and with velocity data approximated by some other method; then, at given time intervals, to replace the computed values of T by the "correct" ones, i.e., to "update" T. However, in most experiments with this method, "correct" values were taken from a "control run", because of the unavailability of real data with the required uniform distribution.

The difficulties encountered with this technique and variations thereof (Morel et al., 1971, Williamson and Kasahara, 1971, Mesinger, 1972) were the following: (a) decrease of the initial root-mean-square error in (u, v) to a non-zero asymptotic value, (b) achievement of this decrease within periods not shorter than two days, and (c) slower decrease or actual increase of the rms error in (u, v) when using temperature data from a control run of a different GCM or real data. Better results were obtained recently by certain improvements of the updating technique such as local balancing (Stone et al., 1973, Kistler and McPherson, 1975) and relaxation (Davies and Turner, 1975), but the basic problems remained. The numerous and varied contributions in this field have been reviewed by Kasahara (1972), Jastrow and Halem (1973), and most recently and completely by Bengtsson (1975).

In view of the difficulties encountered by the non-synoptic techniques (updating, four-dimensional data assimilation) as well as by the traditional synoptic techniques (approximate-balance equations, objective analysis), a different approach to the initialization problem is attempted here. Indeed, if the primitive equations are a good mathematical model for large-scale phenomena in the atmosphere, then there ought to exist initial states $(u, v, \rho, T)\,|_{t=t_0}$ which characterize solutions of the model distinguished by small-amplitude IGW components; and a similar statement has to hold for the numerical models approximating the mathematical one. One should then be able to describe such initial states by equations derived from the model itself, without any approximations.

On the other hand, it should be possible to obtain in the same context diagnostic equations which determine the horizontal velocity field (u, v) from the mass field and a number of its time derivatives. This would seem to be the rigorous mathematical expression of the ideas behind the updating techniques. These two points of view are concerned with the questions of accuracy and of completeness of the initial data respectively; their combination leads to the search for equations, derived from those of the model, in which no time derivatives of the velocity components u, v appear. From these equations u, v can then be determined at any time $t=t_0$, given the thermodynamic variables of state T, p, ρ appearing in the equations, and as many of their time derivatives as required. These time derivatives can be considered to play here the role of the time history of the variables concerned, if we think that t_1 has been let to tend to t_o in an updating process.

This approach is also related to the more elaborate diagnostic equations proposed by Fjørtoft (1962), Hinkelmann (1962) and Hollmann (1966) among others as an improvement on the classical balance equation. The latter is essentially based on the assumption that $d(u_x+v_y)/dt$ is negligible with respect to other terms in the equation and is therefore set to zero[2]; the later, more sophisticated diagnostic equations referred to were all based on assumptions about the vanishing of higher derivatives of the horizontal velocity field or of its divergence, such as $d^2(u,v)/dt^2$ or $d^2(u_x+v_y)/dt^2$, as well as combinations thereof and of the original assumption that u_x+v_y itself or $d(u_x+v_y)/dt$ are zero.[3] Hence these diagnostic equations were in a sense similar to the different closure approximations of turbulence theory.

We indicated already how our procedure can be thought of as the limiting result of updating over a time interval which tends to zero. In a similar vein our diagnostic equations can be viewed heuristically as the limit of a sequence of closure-type equations when the order of the derivatives assumed to be zero tends to infinity.

The advantage of a zero-length updating interval is that the actual solution does not change during the updating, so that the asymptotic error in updating should be reduced to zero. Similarly, the advantage of closure at infinity should be the lack of an initialization shock, i. e., the presence in the time-dependent solution of only the amount of inertia-gravity waves actually attributable to the solution, frozen as it were at

---

[2] Here x, y are cartesian horizontal coordinates and d/dt is horizontal material derivative, $d/dt=\partial/\partial t+u\partial/\partial x+v\partial/\partial y$

[3] Miyakoda and Moyer (1968) proposed a dynamical implementation of such a synoptic condition.

the initial instant, without spurious amplification due to incompatibility between the measured and the derived quantities in the inital state. Indeed, since our diagnostic equations are derived directly from the equations governing the time evolution of the system, without additional assumptions, the dynamical and the thermodynamical quantities in the initial state are compatible with each other with respect to their evolution in time exactly, and not only approximately so.

In Section 2 we derive such diagnostic equations for the shallow-fluid equations linearized around a state of rest and we give bounds on the errors in the derived velocity field in terms of the errors in the measured geopotential field. In Section 3 we derive the diagnostic equations for the full nonlinear shallow-fluid equations and we analyze the type of the equations obtained; a simple ellipticity condition, which does not depend on the space derivatives of u,v, is given. We also point out the relationship our results bear to the classical balance equation and its well-known type analysis. The results of Section 3 are immediately applicable to the primitive equations in isentropic coordinates, which are useful in short-term meso-scale prediction. In Section 4 we show how to extend the results of Section 3 to the Euler equations and to the Navier-Stokes equations, and then we address ourselves to the primitive equations in pressure coordinates. The type analysis of the diagnostic equations obtained is more difficult because of the equations being more complicated, but mainly because of three, rather than two space variables being involved. Still the ellipticity condition can be formulated quite generally, but no simple form of this condition, which does not involve space derivatives of u, v, could be found.

2. Diagnostic Equations for the Linearized Shallow-Fluid Equation Model

The simplest model whose solutions exhibit behavior similar to the large-scale motions of the atmosphere is the model governed by the linearized shallow-fluid equations. In a rotating Cartesian x,y - coordinate system these equations are

$$\text{(1a)} \qquad u_t + \phi_x - fv = 0,$$

$$\text{(1b)} \qquad v_t + \phi_y + fu = 0,$$

$$\text{(1c)} \qquad \phi_t + \Phi(u_x+v_y) = 0,$$

when the linearization is performed around a state of rest. Here u, v are the velocity components in the x,y directions respectively and f is the Coriolis parameter which determines the influence of the rotation on the solutions. For convenience the geopotential ϕ is introduced instead of the height of the free surface h by

$$\phi = gh,$$

with g the acceleration of gravity; the equilibrium value of the geopotential is

$$\Phi = \text{const.}$$

We note that in this model the assumption that the mass field is known is equivalent simply to assuming that we know $\phi = \phi(x,y,t)$, and hence ϕ_t, ϕ_{tt} as well.

Our purpose with respect to system (1) is to obtain two equations in which u_t, v_t do not appear, although ϕ_t, ϕ_{tt} may. Clearly (1c) itself is such an equation, and only one additional equation satisfying the requirements has to be derived. We proceed to do so by differentiating (1c) with respect to t, (1a) and (1b) with respect to x,y, and obtain

(2a) $$u_{tx} + \phi_{xx} - fv_x = 0,$$

(2b) $$v_{ty} + \phi_{yy} + fu_y = 0,$$

(2c) $$\phi_{tt} + \Phi(u_{xt} + v_{yt}) = 0,$$

where we assume for simplicity that f = const., as we shall throughout this section. Substituting u_{xt} and v_{yt} from (2a), (2b) into (2c) yields

(3) $$\Phi f(v_x - u_y) - \Phi(\phi_{xx} + \phi_{yy}) + \phi_{tt} = 0$$

Hence (3) together with (1c) form a system of two first-order partial differential equations for u, v,

(4a) $$u_x + v_y = -\phi_t/\Phi,$$

(4b) $$u_y - v_x = (\phi_{tt} - \Phi\Delta\phi)/\Phi f,$$

where Δ is the two-dimensional Laplacian operator,

$$\Delta\phi = \phi_{xx} + \phi_{yy}.$$

Thus (4) is the required set of diagnostic equations for the model whose time evolution is given by the prognostic equations (1). We notice that the instantaneous (synoptic) determination of u, v at $t = t_0$ from (4) actually requires only a knowledge of ϕ_t, ϕ_{tt} at $t = t_0$, rather than the whole time history of ϕ from some $t = t_1 < t_0$ until $t = t_0$: the equivalent of the time history in this formulation would be a knowledge of all the time derivatives of ϕ.

System (4) for the functions u, v is elliptic, i.e., it has no real characteristics. In fact it is just a set of inhomogeneous Cauchy-Riemann equations for the functions v, u; by cross-differentiation these equations lead to a Poisson equation for either u or v. A well-posed problem for (4) would be the Dirichlet problem. This means prescribing u say on a closed contour ∂D ; then v is determined also up to an additive constant, the value of which can be given by prescribing v at some point on the contour ∂D or in its interior, D (e.g., Miranda, 1970, p. 265 ff.). Thus, within the framework of model (1), solving the Dirichlet problem for system (4) solves the completeness-of-data problem. We turn now to discussing the accuracy problem for the model at hand.

It is well known (Hinkelmann, 1951, Morel et al., 1971, Williamson and Dickinson, 1972) that system (1) has three independent plane-wave solutions, one corresponding to slow Rossby waves, the other two to inertia-gravity waves propagating in opposite directions. In the case we treat, in which the unperturbed velocity is zero[1], the Rossby mode is stationary and the IGWs have phase velocity

$$c = \pm(k^2\Phi+f^2)^{\frac{1}{2}}/k,$$

where $\underset{\sim}{k} = (k_1, k_2)$ is the wave vector and

$$k^2 = k_1^2 + k_2^2.$$

Any solution of (1) can be represented by a series expansion with respect to $\underset{\sim}{k}$ in these plane waves.

Let the solution vector $w = (\phi, u, v)$ of (1) be decomposed into its Rossby component $\bar{w}$ and its IGW component w',

$$(5) \qquad w = \bar{w} + w',$$

with $\bar{w}$ stationary, that is, $\bar{w}_t \equiv 0$. The Rossby component $\bar{w}$ satisfies

$$(1') \qquad f\bar{u} = -\bar{\phi}_y, \quad f\bar{v} = \bar{\phi}_x,$$

i.e., it is geostrophically balanced; it also satisfies (4) with $\partial/\partial t \equiv 0$.

[1] Notice that linearization around a solution with velocity (U, V) satisfying $fU = -\Phi_y$, $fV = \Phi_x$, $\Phi \neq$ const., would still allow us to eliminate the time derivatives of u, v by cross-differentiation in (1). However second space derivatives of u, v would then appear in the equation that corresponds to (4b), making the analysis more difficult. Similarly if f depended on y, say, a term $f_y u/f$ would appear on the right-hand side of (4b), again complicating the analysis.

We have already seen that u, v can be determined in a domain D by solving (4) when ϕ, ϕ_t, ϕ_{tt} are given in D and u, v are given on its boundary ∂D. Now we want to show that actually $\bar{u}, \bar{v}$ can be determined with prescribed accuracy, even when the data ϕ, ϕ_t, ϕ_{tt} in D and u, v on ∂D do not correspond to a solution $\bar{w}$ in geostrophic balance. We start by pointing out the relationship between geostrophicity throughout D and geostrophicity on the boundary ∂D.

Notice first that the most general solution of (4) which corresponds to a stationary solution of (1) will satisfy

(6a) $$u_x + v_y = 0$$

(6b) $$v_x - u_y = \Delta\chi,$$

where $\chi = \phi/f$ is a known function. The solution of (6a) is

(7) $$u = -\psi_y, \quad v = \psi_x,$$

with ψ an arbitrary, twice continuously differentiable function. By (6b) ψ has to satisfy

(8) $$\Delta(\psi - \chi) = 0$$

Hence, if ψ satisfies the Dirichlet boundary condition

$$\psi = \chi \quad \text{on} \quad \partial D,$$

then

$$\psi = \chi \quad \text{in} \quad D.$$

Similarly, if ψ satisfies the Neumann boundary condition

(9) $$\partial_n \psi = \partial_n \chi \quad \text{on} \quad \partial D,$$

where

$$\partial_n = n \cdot \nabla$$

and n is the unit (outer) normal to ∂D, ∇ the gradient operator, then

(10) $\qquad \psi = \chi + \text{const.} \qquad \text{in} \quad D.$

But we have from (7) that

$$\partial_n \psi = n \cdot (v, -u)$$

so that (9) becomes

(9') $\qquad n \cdot (v - \chi_x, -u - \chi_y) = 0 \qquad \text{on} \quad D.$

The simplest particular instance of (9') is

(9") $\qquad u = -\chi_y, \; v = \chi_x \qquad \text{on} \quad \partial D;$

in fact it is the only one. Indeed, we saw that equation (8) with boundary condition (9) has the solution (10); furthermore, (10) and (7) imply

(11) $\qquad u = -\chi_y, \; v = \chi_x \qquad \text{in} \quad D.$

Hence we conclude, by a continuity argument, that (11) implies (9"), i.e., (9) implies (9").

Summing up our discussion of system (6), it is clear that its most general solution, whether it satisfy Dirichlet or Neumann boundary conditions, is geostrophic in D if and only if it is geostrophic on ∂D. Let us now return to the decomposition (5), where in the sequel w will be taken to stand for a solution of (1) corresponding to initial data obtained by solving the diagnostic system (4). Refine the decomposition for this w as

(5') $\qquad w = \bar{w} + w_1 + w_2,$

which is equivalent to writing

$$w' = w_1 + w_2.$$

Here $\bar{w}$ is the geostrophic or Rossby component of w as before, w_1 is the deviation of w from $\bar{w}$ due to departure from geostrophicity in the boundary conditions of (4) and w_2 is the deviation due to departure from geostrophicity in the right-hand side of (4). Further on we shall make these definitions precise, and in the process show how to obtain $\bar{w}$ in (5').

We start by observing that, if a function ρ can be identified as the geopotential ϕ in a solution $w = (\phi,u,v)$ of (1), $\rho = \phi$, then ρ has to satisfy

$$(12) \quad \partial/\partial t\{\partial^2/\partial t^2 - \Phi(\partial^2/\partial x^2+\partial^2/\partial y^2) + f^2\}\rho = 0$$

(e.g., Courant and Hilbert, 1962, pp. 14-15)[2]. Such a function ρ will be uniquely determined by (12), provided initial data ρ, ρ_t, ρ_{tt} are prescribed at $t = t_0$. The general solution of (12), like that of (1), can be expressed as an infinite sum of plane waves of three types, one Rossby mode and two inertia-gravity modes.

Now let ϕ, ϕ_t, ϕ_{tt} be given in D at $t = t_0$, with $\phi_t \neq 0$, $\phi_{tt} \neq 0$ in general. Solve (12) with

$$\partial^k/\partial t^k \rho = \partial^k/\partial t^k \phi, \; k = 0,1,2, \; \text{at} \; t=t_0.$$

Represent the solution

$$\rho = \phi(x,y,t)$$

by its plane-wave decomposition, and eliminate the IGW terms by setting their coefficients equal to zero. The new function with only Rossby components in its plane-wave expansion is $\bar{\phi}$; in particular we have

$$\bar{\phi}_t = 0, \; \bar{\phi}_{tt} = 0 \; \text{at} \; t = t_0,$$

since $\bar{\phi}_t \equiv 0$. The physical interpretation is that $\phi - \bar{\phi}$ results from measurement errors of $\partial^k/\partial t^k \phi$ at $t = t_0$. In the final section we shall discuss how to obtain ϕ_t, ϕ_{tt} from actual observations in an optimal manner.

More generally, we might desire to reduce IGW terms to a realistic size, rather than eliminate them completely. It suffices then to multiply their coefficients by suitably small numbers, instead of setting them to zero. We shall call ϕ modified by such a filtering procedure $\tilde{\phi}$; $\bar{\phi}$ is thus a particular case of $\tilde{\phi}$.

---

[2] Here ρ is <u>not</u> density, which does not appear in this section.

We are now ready to define w_1 and w_2 precisely and to compute $\bar{w}$. Clearly $\bar{w}$ is the solution $(\bar{\phi},\bar{u},\bar{v})$ of (1) with $\bar{\phi}$ obtained by the procedure outlined above and with

$$\bar{u} = -\bar{\phi}_y/f, \quad \bar{v} = \bar{\phi}_x/f \quad \text{in } D.$$

Indeed, we have shown in the analysis of system (6) that its solutions, i.e., the solutions of (4) for which $\phi_t = \phi_{tt} = 0$, are geostrophic in D if and only if they are so on ∂D. Notice that $\bar{w}$ is different from the solution $\hat{w}$, say, of (1) with initial data

$$\hat{w}|_{t=t_0} = (\phi,-\phi_y/f,\phi_x/f)|_{t=t_0}.$$

In fact $\hat{w}$ is not stationary and will not remain geostrophic unless $\phi_t = \phi_{tt} = 0$.

According to our heuristic description of w_1 and of w_2 it also becomes clear now what their definitions should be:

w_1 is the solution of (1) with initial data $(0,u_1,v_1)|_{t=t_0}$, where $(u_1,v_1)|_{t=t_0}$ is the solution of (6) with homogeneous right-hand side (RHS), $\chi\equiv 0$, and with boundary conditions

$$(13) \qquad u_1 = u + \bar{\phi}_y/f, \; v_1 = v - \bar{\phi}_x/f \qquad \text{on } \partial D,$$

u,v, being the values actually measured on ∂D;

w_2 is the solution of (1) with initial data $(\phi-\bar{\phi},u_2,v_2)|_{t=t_0}$, where $\phi|_{t=t_0}$ is prescribed and $(u_2,v_2)|_{t=t_0}$ is the solution of (4) with $\phi-\bar{\phi}$ instead of ϕ on the RHS and with homogeneous boundary conditions

$$u_2=0, \; v_2=0 \quad \text{on } \partial D.$$

As we saw in Section 1, the situation which more closely resembles that of the atmosphere, and which we expect therefore in a primitive-equation model, is the presence of small-amplitude IGWs rather than their total absence. Hence we might not want to use $\bar{w}$, which contains no IGW components at all, but a suitably modified $\tilde{w}$ with realistically small IGW components. It is in discussing such a $\tilde{w}$ that the definition of w_1 and of w_2 will prove useful.

The first step in obtaining $\tilde{w}$, that of modifying ϕ, has already been described in our discussion of equation (12); it essentially concerns the reduction of the size of w_2. The second step is to modify the boundary conditions for $(u_1,v_1)|_{t=t_0}$ so that, for a given $\varepsilon>0$, the inequalities

$$(13') \qquad |u_1|\leq\varepsilon, |v_1|\leq\varepsilon \qquad \text{on } \partial D$$

hold at $t = t_0$. Call the solution of (6) with $\chi\equiv0$ and boundary conditions thus modified $(\overset{o}{u}_1,\overset{o}{v}_1)$. By the maximum principle for the Laplace equation (Courant and Hilbert, 1962) the inequalities (13') imply

$$(14) \qquad |\overset{o}{u}_1|\leq\varepsilon, \ |\overset{o}{v}_1| \leq\varepsilon \quad \text{in } D.$$

This is a limitation on the size of $\tilde{w}_1 = (0,\tilde{u}_1,\tilde{v}_1)$ at $t=t_0$, where $(\tilde{u}_1,\tilde{v}_1)|_{t=t_0} = (\overset{o}{u}_1,\overset{o}{v}_1)$.

Denote by $(\overset{o}{u}_2, \overset{o}{v}_2)$ the solution of (4) with $\tilde{\phi}-\bar{\phi}$ instead of ϕ on the RHS and with homogeneous boundary conditions. The theory of elliptic PDEs (Miranda, 1970) provides us with certain bounds on $(\overset{o}{u}_2,\overset{o}{v}_2)$ in terms of bounds imposed on $\tilde{\phi}-\bar{\phi}$ at $t=t_0$; these so-called error estimates are similar in character to and are partly derived from the maximum principle. Furthermore, the theory of hyperbolic PDEs (Courant and Hilbert, 1962), applied to system (1), states that under suitable technical restrictions there are bounds on the size of $\tilde{w}_1$ and $\tilde{w}_2$ at any time t in terms of their size at the initial time $t=t_0$ (energy inequalities). Thus, ultimately, there are bounds on $\tilde{w}-\bar{w}$ at any time t in terms of the bounds at $t=t_0$ on $\tilde{\phi}-\bar{\phi}$, $\tilde{\phi}_t$, $\tilde{\phi}_{tt}$ in D and on $\tilde{u}+\bar{\phi}_y/f$, $\tilde{v}-\bar{\phi}_x/f$ on ∂D.

This completes the discussion of the accuracy question for our initialization by compatible balancing in the case of a linearized shallow-fluid model.

3. Diagnostic Equations for the Nonlinear Shallow-Fluid Equation Model

The solutions of this model have more features in common with the large-scale motions of the atmosphere than those of the model treated in the previous section. The model also exhibits some of the behavior associated with the nonlinearity of the primitive equations; it is considered useful in many analytical and numerical investigations which aim at applications to the primitive equations. The governing equations are:

$$\text{(1a)} \qquad u_t + uu_x + vu_y + \phi_x - fv = 0 \quad ,$$

$$\text{(1b)} \qquad v_t + uv_x + vv_y + \phi_y + fu = 0 \quad ,$$

$$\text{(1)} \qquad \phi_t + u\phi_x + v\phi_y + \phi(u_x+v_y) = 0 \quad ,$$

where the notation is that of Section 2[1]. However, for the sake of completeness, we shall let f , in this section only, be a function of y ; not to include x-dependence as well reflects the customary convention of considering the x-axis to be locally tangent to a circle of latitude, the y-axis to a circle of longitude.

Again we want to derive two equations for the velocity components u, v in which their time derivatives do not appear, although time derivatives of the geopotential ϕ may be present; our standing assumption is that ϕ is known together with its time derivatives. Equation (1c) satisfies these requirements, as (2.1c) did before, and one more such equation has to be found. It was shown in Ghil (1973) that the straightforward generalization of the procedure used in the previous section to obtain such an equation fails. A slight modification thereof, however, works, as we shall show presently. The modification is based on the idea that, in a nonlinear model, the material derivative d/dt is the analog of the partial derivative $\partial/\partial t$ in a linearized model.

For notational convenience we introduce $\Phi = \log\phi$, which is justified

---

[1]We shall start the numbering of equations afresh in each section, and make cross-references to equations in a different section by prefixing the section number, e.g., the linearized shallow-fluid equations will be referred to hereafter as system (2.1).

since ϕ is strictly positive,[2] and the divergence

$$\delta = u_x + v_y \quad .$$

In this notation (1c) becomes

$$(1c') \qquad \delta = - d\Phi/dt \quad ,$$

where d/dt is the two-dimensional material derivative,

$$d/dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y.$$

Differentiating (1a) with respect to x, (1b) with respect to y and adding we obtain the familiar divergence equation, which we write as

$$(2) \qquad - d\delta/dt = u_x^2 + 2u_y v_x + v_y^2 + f(u_y - v_x) + f_y u + \Delta\phi \; .$$

But from (1c') we can express $d\delta/dt$ in (2) also as

$$(3) \qquad - d\delta/dt = d\Phi_t/dt + ud\Phi_x/dt + vd\Phi_y/dt + \Phi_x du/dt + \Phi_y dv/dt \; ,$$

where we use the fact that the material derivative d/dt obeys the product rule for two arbitrary scalar functions G,H,

$$d(GH)/dt = HdG/dt + GdH/dt \quad .$$

Expressing du/dt, dv/dt with the aid of (1a), (1b), equation (3) becomes

$$(3') \qquad - d\delta/dt = \Phi_x(fv - \phi_x) + \Phi_y(-fu - \phi_y) + d\Phi_t/dt + ud\Phi_x/dt + vd\Phi_y/dt \; ,$$

which we can now combine with (2) to yield the desired second diagnostic equation.

We thus obtain the diagnostic system

$$(4a) \qquad u_x + v_y = - \Phi_x u - \Phi_y v - \Phi_t \quad ,$$

$$(4b) \qquad u_x^2 + 2u_y v_x + v_y^2 + f(u_y - v_x) = - f_y u + f(\Phi_x v - \Phi_y u) + d\Phi_t/dt + ud\Phi_x/dt + vd\Phi_y/dt - (\phi_x^2 + \phi_y^2)/\phi - \Delta\phi \quad ,$$

[2]The Φ just defined bears no relationship whatsoever to Φ = const. of Section 2.

for the model whose time evolution is given by the prognostic system (1). The right-hand side of (4) contains u,v, but only in non-differentiated form. In the case in which

$$\delta = d\delta/dt = 0$$

system (4) is easily seen to be equivalent to the classical balance equation

$$(5) \qquad 2(\psi_{xy}^{2} - \psi_{xx}\psi_{yy}) - f\Delta\psi = f_y\psi_y - \Delta\phi \; ,$$

where ψ is the stream function,

$$(6) \qquad u = -\psi_y \quad , \quad v = \psi_x \quad .$$

Equation (5) can also be interpreted as determining the non-divergent part of the horizontal velocity field in a primitive-equation model, where the non-rotational part of the field is much smaller and can be neglected, along with other terms in the horizontal divergence equation of such a model (Haltiner, 1971, p.60).

Notice, however, that (5) fails to yield a good approximation for the horizontal wind field in low latitudes, where some of the assumptions made in deriving it (especially quasi-geostrophicity and small divergence) are not justified any longer. There have been different suggestions on how to circumvent this problem, such as taking ψ to be given in the topics and solving (5) for ϕ there (Houghton and Washington, 1969). Still we believe it is a definite advantage of (4) that no assumptions whatsoever were made in deriving it from (1); therefore its solution $(u,v)|_{t=t_0}$ should provide an initial state that is entirely compatible (in the sense described in Section 1) with the time evolution of the solution (ϕ,u,v) of (1), independently of latitude.

In order to solve (4) for (u,v) we still need to determine appropriate boundary conditions. No general theory of well-posedness exists for such nonlinear first-order systems[3], but the known results for linear systems suggest that the nature of appropriate boundary conditions depends on the type of the system. We saw in the discussion of (2.4) that the Dirichlet problem is well-posed for linear elliptic systems. For hyperbolic systems the number and nature of the data to be prescribed along the boundary varies from one piece of boundary to another, according to the geometry of the domain and of the characteristics (Kreiss and Oliger, 1973, pp. 64-70). Finally in connection with systems of mixed type (i.e., systems which have different type in different parts of the domain considered) relatively little is known; however, some advances have been recently made for linear and quasi-linear systems encountered in the field of transonic flows (Jameson, 1975): we shall have to return to this topic at a later point of our discussion.

Thus our next task is to determine the type of system (4). We recall first the familiar theory for the single nonlinear equation of second order

$$(7) \quad F(\psi) \equiv E(\psi_{xx}\psi_{yy}-\psi_{xy}^2) + A\psi_{xx} + 2B\psi_{xy} + C\psi_{yy} + D = 0 ,$$

where the coefficients A,B,C,D,E are (continuous) functions of x,y,ψ,ψ_x,ψ_y. If $E \neq 0$, then (7) is a Monge-Ampère equation; it was most recently discussed in the meteorological literature by Houghton (1968). Clearly (5) is a special instance of (7) with

$$(8) \qquad E = 2, \quad A = C = f , \quad B = 0 , \quad D = f_y\psi_y - \Delta\phi .$$

The meaning of characteristics for such a nonlinear equation is explained by Courant and Hilbert (1962, pp. 418-421) in terms of the solvability of the Cauchy problem. Given any second-order nonlinear PDE,

$$(7') \qquad F(x,y,\psi,\psi_x,\psi_y,\psi_{xx},\psi_{xy},\psi_{yy}) = 0 ,$$

its characteristic lines

$$\Psi(x,y) = \text{const.}$$

---

[3] A system like (4) is in general not equivalent to a single second-order equation (Courant and Hilbert, 1962, pp. 12-14). Furthermore, to treat a higher-order equation is not necessarily more advantageous than to treat a first-order system. In any case, we have not found such an equivalent single equation.

are given by

$$(9) \qquad F_{\psi_{xx}}\Psi_x^2 + F_{\psi_{xy}}\Psi_x\Psi_y + F_{\psi_{yy}}\Psi_y^2 = 0 \quad .$$

Ellipticity is defined as the nonexistence of real characteristics; hence the condition for (7') to be elliptic is

$$(10) \qquad 4\Delta^2 \equiv F^2_{\psi_{xy}} - 4F_{\psi_{xx}}F_{\psi_{yy}} < 0 \quad .$$

The interesting feature of (7) is that (Courant and Hilbert, 1962, pp. 495-499) the discriminant Δ^2 is independent of the highest derivatives,

$$(10') \qquad \begin{aligned} \Delta^2 &\equiv (B-E\psi_{xy})^2 - (A+E\psi_{yy})(C+E\psi_{xx}) \\ &= B^2 - AC - E\{2B\psi_{xy} + A\psi_{xx} + C\psi_{yy} + E(\psi_{xx}\psi_{yy}-\psi_{xy}^2)\} \\ &= B^2 - AC + DE \quad . \end{aligned}$$

Combining (10) and (10') yields the ellipticity condition

$$(11) \qquad B^2 - AC + DE < 0$$

for (7); by (8) this becomes

$$(11') \qquad 2(\Delta\phi - f_y\psi_y) + f^2 > 0 \quad ,$$

which is the well-known ellipticity condition for equation (5). We conclude this discussion of (7) by mentioning that, given (11), one can show essential uniqueness (viz.,existence of at most two solutions, each with different geometrical properties: Rellich, 1932, Courant and Hilbert, 1962, pp. 324-326), and also existence (Pogorelov, 1964, pp. 88-92) of solutions satisfying boundary conditions of Dirichlet type.

After these remarks on (7), we are ready to proceed with the discussion of the first-order nonlinear system (4), which we write for convenience as

$$(12a) \qquad D(w_x, w_y) = d(x,y,w) \quad ,$$

$$(12b) \qquad E(w_x, w_y) = e(x,y,w) \quad ,$$

where now $w = (u,v)$. In order to stress further the analogy with (7'), (12) can be written even more concisely as

$$(12') \qquad F(x,y,w,w_x,w_y) = 0 \quad ,$$

where $F = (D-d, E-e)$. The generalization of the concept of characteristics to (12), starting either from (7') or from quasi-linear systems (Courant and Hilbert, 1962, pp. 170-173 and 424-427) leads to the characteristic equation

$$(13) \qquad 0 = \det \{(\partial F/\partial w_x)\Psi_x + (\partial F/\partial w_y)\,\Psi_y\}$$

$$\equiv \det \left\{ \begin{pmatrix} D_{u_x} & E_{u_x} \\ D_{v_x} & E_{v_x} \end{pmatrix} \Psi_x + \begin{pmatrix} D_{u_y} & E_{u_y} \\ D_{v_y} & E_{v_y} \end{pmatrix} \Psi_y \right\}$$

After some computations, (13) can be written as

$$(13') \qquad (2u_y - f)\Psi_x^2 + 2(v_y - u_x)\Psi_x\Psi_y - (2v_x + f)\Psi_y^2 = 0 \quad .$$

For (13') not to have real solutions Ψ_x/Ψ_y (or Ψ_y/Ψ_x) the discriminant Δ^2 ,

$$(14) \qquad \Delta^2 = (v_y - u_x)^2 + (2v_x + f)(2u_y - f) \ ,$$

has to be negative, i.e., the ellipticity condition for (12) is

$$(15) \qquad \Delta^2 < 0 \ .$$

But, expanding and rearranging (14), we can put it into the form

$$(14') \qquad \Delta^2 = 2e - d^2 - f^2 \quad ,$$

where we use

$$-\,2u_x v_y = -\,(u_x + v_y)^2 + (u_x^2 + v_y^2) \ .$$

This finally yields the ellipticity condition for (4) as

$$(15') \qquad 2e - d^2 - f^2 < 0,$$

which is the equivalent of (11') for (5). Indeed, the assumption

$$\delta = d\delta/dt = 0 \ ,$$

i.e.,

$$d = 0, \ e = - f_y u - \Delta\phi \equiv e_1 \ ,$$

reduces (15') to the familiar ellipticity condition for the balance equation, exactly as it reduced (4) to (5). Furthermore, condition (15') does not contain any derivatives of u,v, just as (11') did not contain the second derivatives of ψ . This is of particular importance when actually trying to solve (4) numerically, since difference quotients of a numerical solution are in general poor approximations to the derivatives of the exact solution, even though the solution of the difference equations might be a good approximation to the solution of the differential equations.

In the case in which (15') is satisfied throughout the domain of interest we can therefore expect a boundary-value problem such as that discussed for (2.4) to be well-posed and to lead to at most two solutions (cf. our discussion of (5))[4]. By analogy to Rellich's results (Rellich, 1932, Proposition 2) for (7), these two solutions would be easily distinguishable from each other and only one of them could be physically significant under given conditions (in the other one cyclones and anticyclones would be interchanged).

The major difficulty one might encounter in solving (4) is that of change of type. It is well known from experience with (5) that for certain geopotential data (11') is not satisfied, especially in regions of strong anticyclonic activity. The same is to be expected of (4), since, in terms of a scale analysis of (15'), e_1 clearly dominates $(e-e_1) - d^2/2$ over most of the earth.

The customary approach in solving (5) numerically when the data indicate hyperbolicity has been to modify the data so that (11') is satisfied (Shuman, 1957,

[4] Heuristically two solutions are to be expected in fact because of (4b) being quadratic in the derivatives of u, v.

Krishnamurti, 1968, Paegle and Paegle, 1974). This approach guaranteed that the known mathematical theory of (7) be applicable and that the boundary-value problem be well-posed; furthermore, elliptic difference operators could be used to approximate (5) throughout the domain, which was convenient and efficient. However, the numerical solutions thus obtained differed from observations considerably, at least in the parts of the domain where the data had to be modified, and they could not lead to over-all satisfactory results. We suggest a different line of attack, motivated by recent advances in dealing with equations of mixed type.

In the area of transonic gas dynamics equations of mixed type arise that are linear (the Tricomi equation, Bers, 1958, p. 22) or quasi-linear, i.e., linear in the highest derivatives (the Chaplygin equation, id., p. 14). In this area a great deal of progress has been made lately in obtaining theoretical results (ibid., Ch. 5 and Ch. 6, Garabedian, 1964, Sec. 12.1) as well as numerical solutions (Murman and Cole, 1971, Jameson, 1975). In particular, it has been shown that solutions of the Chaplygin equation with interior regions of hyperbolicity (of supersonic flow) exist and can be computed efficiently by utilizing elliptic difference operators in the subsonic region and hyperbolic difference operators in the supersonic region. This would seem to encourage investigations in which the regions that are hyperbolic for system (4) (or for equation (5) for that matter) are accounted for, rather than ignored. In this regard it seems that physical intuition would support the well-posedness of the boundary-value problem for (4), including the possibility of interior regions of hyperbolicity. The heuristic argument relies on the compatibility of (4) with (1), putting at least the well-posedness of the two systems on the same footing. Although the boundary conditions under which the initial-and-boundary value problem for (1) is well-posed are not known rigorously (they are for (2.1): Elvius and Sundström, 1973), numerical experience and physical insight suggest that such conditions exist. Unfortunately the same argument cannot be brought to bear on (5), since it is not compatible with (1) in the sense discussed in the Introduction; neglecting certain physically small terms in an equation might have far-reaching mathematical consequences, as is well-known from the theory of boundary layers and of other singular perturbation phenomena (Cole, 1968).

We conclude this section by pointing out that its results can be applied to the primitive equations in isentropic coordinates (Eliassen and Kleinschmidt, 1957, pp. 26-27),

$$u_t + uu_x + vu_y + M_x - fv = 0 \quad , \tag{16a}$$

$$v_t + uv_x + vv_y + M_y + fu = 0 \quad , \tag{16b}$$

$$\pi_t + u\pi_x + v\pi_y + \pi(u_x+v_y) = 0 \quad . \tag{16c}$$

where M is the Montgomery potential,

$$M = \phi + c_p T \quad ,$$

with c_p the specific heat at constant pressure. Here π is given by

$$\pi = \partial p/\partial\theta \ ,$$

where θ is the potential temperature,

$$\theta = (p_o/p)^{1-1/\kappa}\ T \ ,$$

and

$$p_o = 1000\text{mbar}, \quad \kappa = c_p/c_v \ ,$$

with c_v the specific heat at constant volume (id., p. 3). Indeed, (16) is entirely similar to (1) with ϕ replaced by M in (1a) and (1b) and by π in (1c); moreover knowledge of the mass field T, p, ρ implies knowledge of M

and π in (16). Thus, letting now

$$\Phi = \log\pi ,$$

we obtain, in a manner entirely analogous to (4), the diagnostic system

$$u_x + v_y = - \Phi_x u - \Phi_y v - \Phi_t \quad , \tag{17a}$$

$$\begin{aligned} u_x^2 + 2u_y v_x + v_y^2 + f(u_y - v_x) = &- f_y u + f(\Phi_x v - \Phi_y u) \\ &+ d\Phi_t/dt + ud\Phi_x/dt + vd\Phi_y/dt \\ &- (\Phi_x M_x + \Phi_y M_y) - \Delta M \ , \end{aligned} \tag{17b}$$

to which the same discussion applies. As we indicated already, the study of (1) and hence of (4) has research value with regard to properties of the primitive equations in natural or in pressure coordinates. The study of (16) and hence of (17) has even more immediate practical value, because of the usefulness of (16) in meso-scale (short-range) weather prediction (Bleck, 1974).

4. Diagnostic Equations for the Primitive-Equation Model

This model is currently believed to be the most useful in describing large-scale motions of the atmosphere and different forms thereof are used in operational weather forecasting, as well as in research-oriented general circulation models. It is closely related to the Euler equations of fluid dynamics,

$$\text{(1a)}\quad \frac{du}{dt} = -\frac{1}{\rho} p_x + fv,$$

$$\text{(1b)}\quad \frac{dv}{dt} = -\frac{1}{\rho} p_y - fu,$$

$$\text{(1c)}\quad \frac{dw}{dt} = -\frac{1}{\rho} p_z - g,$$

$$\text{(1d)}\quad \frac{d\rho}{dt} = -\rho(u_x+v_y+w_z),$$

$$\text{(1e)}\quad \frac{d\theta}{dt} = 0,$$

where d/dt is now the three-dimensional material derivative,

$$d/dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y + w\partial/\partial z,$$

with z the vertical coordinate and w the corresponding velocity component.

For the sake of completeness we mention that the procedure of the previous section immediately extends to (1), yielding for it the diagnostic system

$$\text{(2a)}\quad u_x + v_y + w_z = -\Phi_x u - \Phi_y v - \Phi_t,$$

$$\text{(2b)}\quad u_x^2 + v_y^2 + w_z^2 + 2(u_y v_x + v_z w_y + w_x u_z) + f(u_y - v_x)$$
$$= f(\Phi_x v - \Phi_y u)$$
$$+ d\Phi_t/dt + ud\Phi_x/dt + vd\Phi_y/dt + wd\Phi_z/dt$$
$$- \frac{1}{\rho}\Delta p - g\Phi_z,$$

$$\text{(2c)}\quad \theta_x u + \theta_y v + \theta_z w = -\theta_t.$$

Here Δ is the three-dimensional Laplacian,

$$\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2,$$

and

$$\Phi = \log \rho,$$

while f,g are taken as constants for simplicity. For system (2) to be useful in determining (u,v,w) two thermodynamic variables, T and p, say, have to be observable throughout the domain of interest. This is not so unreasonable to require: it is still easier to measure one additional scalar, p, rather than the three velocity components u,v,w. Hence (2) could prove useful in the study of certain small-scale atmospheric phenomena where vertical accelerations are not negligible, as well as in other areas of fluid dynamics where system (1) arises and where temperature and pressure measurements are easier to make than velocity measurements.

The same result holds for the full Navier-Stokes equations, only that the equation corresponding to (2c) will contain derivatives of the velocity components, instead of being purely algebraic in them (since in the equation corresponding to (1e) there is an energy dissipation term depending on velocity derivatives); also the equation corresponding to (2b) will be of order higher than first (second or third according to whether the flow studied is incompressible or compressible), but linear in the highest derivatives. Since we have no immediate interest in these equations, we turn our attention now to the primitive-equation model.

For simplicity we shall consider the primitive equations in pressure coordinates and without dissipation or heating terms,

(3a) $u_x + v_y + \omega_p = 0$

(3b) $u_t + uu_x + vu_y + \omega u_p - fv = -\phi_x,$

(3c) $v_t + uv_x + vv_y + \omega v_p + fu = -\phi_y,$

(3d) $\theta_t + u\theta_x + v\theta_y + \omega\theta_p = 0,$

where θ is potential temperature, ϕ is the geopotential of an isobaric surface,

$$\phi = gz(x,y,p), \; p = \text{const.},$$

and $\omega = dp/dt$ is the vertical velocity in this coordinate system. The results below obtain also for the equations in natural z-coordinates or in σ-coordinates. If heating and dissipation are present, and are either known or depend on the velocity components and their spatial derivatives in a known way, the same comments as for the Navier-Stokes equations apply: the corresponding diagnostic equations will be more complicated than the ones below and possibly of higher order, but instantaneous, synoptic determination of the velocity field from the mass field is still possible.

We have now three velocity components, u, v, ω, to determine and need therefore three diagnostic equations. Obviously (3a) and (3d) are two such equations, and we have to derive a third one. The basic idea will be again to start from a divergence equation and eliminate the material derivative of the divergence using the continuity equation. The procedure is less straightforward in the case of system (3) than in the case of system (1) because of the asymmetric way in which ω appears in (3).

We introduce the horizontal divergence δ,

$$\delta = u_x + v_y,$$

and the material derivative d/dt corresponding to this coordinate system,

$$d/dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y + \omega\partial/\partial p.$$

In this notation the divergence equation obtained from the momentum equations (3b), (3c) is

(4) $$d\delta/dt + u_x^2 + 2u_y v_x + v_y^2 + \omega_x u_p + \omega_y v_p = f(v_x - u_y) - \Delta\phi.$$

From the continuity equation (3a), written as

(3a') $$\delta = -\omega_p,$$

we have immediately

(5) $$-d\delta/dt = d\omega_p/dt.$$

We shall use the energy equation (3d) to eliminate ω,

(3d') $$\omega = -(\theta_x u + \theta_y v + \theta_t)/\theta_p = -\theta' u - \theta'' v - \theta''',$$

with the obvious identifications

$$\theta' = \theta_x/\theta_p, \; \theta'' = \theta_y/\theta_p, \; \theta''' = \theta_t/\theta_p.$$

Thus, from (5) and (3d'), we have

(5') $$d\delta/dt = (d/dt)(\theta' u_p + \theta'' v_p + \theta'_p u + \theta''_p v + \theta'''_p)$$

$$= \theta' du_p/dt + \theta'' dv_p/dt + (d\theta'/dt)\, u_p + (d\theta''/dt) v_p$$

$$+ \theta'_p du/dt + \theta''_p dv/dt + u d\theta'_p/dt + v d\theta''_p/dt + d\theta'''_p/dt.$$

We can also write ω_x and ω_y appearing in (4) with the aid of (3d'). Furthermore, equations (3b) and 3c) yield

$$du_p/dt = (fv - \phi_x)_p - u_x u_p - u_y v_p - u_p \omega_p$$

$$= -u_y v_p - u_p v_y + f v_p - \phi x_p,$$

$$dv_p/dt = (-fu - \phi_y)_p - u_p v_x - v_y v_p - v_p \omega_p$$

$$= -u_p v_x - u_x v_p - f u_p - \phi_{yp},$$

where we used (3a) to express ω_p. Substituting this into (5') and the result into (4) produces the missing diagnostic equation.

Notice that we use (3d) to eliminate the space derivatives of ω, so that we are left with a diagnostic system of two equations for the two horizontal velocity components u, v,

(6a) $$u_x - \theta' u_p + v_y - \theta'' v_p = \theta'_p u + \theta''_p v + \theta'''_p,$$

(6b) $$u_x^2 + 2u_y v_x + v_y^2 - \theta'(u_x u_p + 2u_y v_p + u_p v_y)$$

$$- \theta''(u_x v_p + 2u_p v_x + v_p v_y)$$

$$+ f(u_y - v_x) + \{(\theta'_y - \theta''_x)v + \theta'_t - \theta'''_x - f\theta''\} u_p$$

$$+ \{(\theta''_x - \theta'_y)u + \theta''_t - \theta'''_y + f\theta'\} v_p$$

$$=f(\theta_p''u-\theta_p'v) - ud\theta_p'/dt - vd\theta_p''/dt - d\theta_p''' /dt$$

$$- \Delta\phi + (\theta'\phi_x + \theta''\phi_y)_p.$$

After solving (6) for u,v, we can obtain ω immediately either from (3a) or from (3d). Observe also that in the case in which all p derivatives are zero, $\partial/\partial p \equiv 0$, system (6) reduces to the classical balance equation (3.5).

To solve (6), however, we need to know what a well-posed problem for it is, i.e., which boundary conditions will ensure that the solution exists, is unique and depends continuously on the data (mass-field data and boundary data). The first step in that direction is to determine the type of system (6). This is more difficult than for system (3.4) for two reasons: the first and obvious one is that (6) is more complicated, viz.,more terms appear in the equations; this is actually the less important one. The second reason is that, although (6) has only two dependent variables, u and v, the same as (3.4), it has one additional independent variable, p. It is well known that the analysis of partial differential equations in more than two variables is considerably harder (Courant and Hilbert, 1962, p.551 ff., Garabedian, 1964, p. 175 ff.), since the geometric concepts involved become more complex.

With a slight and obvious change from the notation of Section 3 we rewrite (6) as

(7) $$F(x,y,p,w,w_x,w_y,w_p) = 0.$$

The equation that gives the characteristic surfaces

$$\Psi(x,y,p) = \text{const.}$$

of (7) will then be

$$0 = \det \{(\partial F/\partial w_x)\, \Psi_x + (\partial F/\partial w_y)\, \Psi_y + (\partial F/\partial w_p)\Psi_p\}.$$

Introducing the normal vector ξ of a characteristic surface

$$\xi = (\xi_1,\xi_2,\xi_3) = (\Psi_x,\Psi_y,\Psi_p),$$

the characteristic equation becomes

$$(8) \qquad 0 = \det \begin{pmatrix} D_{u_x}\xi_1 + D_{u_y}\xi_2 + D_{u_p}\xi_3 & D_{v_x}\xi_1 + D_{v_y}\xi_2 + D_{v_p}\xi_3 \\ E_{u_x}\xi_1 + E_{u_y}\xi_2 + E_{u_p}\xi_3 & E_{v_x}\xi_1 + E_{v_y}\xi_2 + E_{v_p}\xi_3 \end{pmatrix}$$

The expression in (8) represents a quadratic form in the three components of ξ, in the same way that the expression in (3.13) was a quadratic form in the two variables Ψ_x, Ψ_y; we can write this form as

$$(9) \qquad Q(\xi) = \sum_{i,j=1}^{3} A_{ij}\xi_i\xi_j.$$

Since we want to consider only real values of ξ, the coefficient matrix $A = (A_{ij})$ of this form can be taken, without loss of generality, to be symmetric,

$$A_{ij} = A_{ji}.$$

The coefficients A_{ij} are certain explicit functions of the known quantities in (6), as well as of u, v and of their derivatives.

The necessary and sufficient condition for system (6) to be elliptic is that the quadratic form (9) be definite (Courant and Hilbert, 1962, pp. 552-556 and pp. 579-581), i.e., that $Q(\xi)$ be different from zero for non-zero ξ. The system is hyperbolic iff (if and only if) there exists a non-singular linear transformation, that is, a transformation given by a matrix (B_{ij}) with non-vanishing determinant, which, when applied to ξ,

$$\xi_i = \sum_{j=1}^{3} B_{ij}\eta_j,$$

will bring Q to the form

$$Q(\eta) = -\eta_1^2 + \eta_2^2 + \eta_3^2.$$

In terms of such a transformation, (6) is elliptic iff a non-singular matrix (B_{ij}) exists for which

$$Q(\eta) = \eta_1^2 + \eta_2^2 + \eta_3^2.$$

In two variables, $\xi_3 \equiv 0$, the only other possibility, when the form $Q(\xi) = Q(\xi_1,\xi_2)$ is really quadratic (rather than linear), consists in it being once degenerate, i.e., reducible to

$$Q(\eta) = \eta_1^2.$$

At a point at which this were the case, system (3.4) would be parabolic; such points, if they existed, would in general form the transition line between a region of ellipticity and one of hyperbolicity: the so-called parabolic line (the sonic line in transonic gas dynamics). In three variables, however, there exist other cases in which (6) is neither elliptic nor hyperbolic. In particular one might have points at which (B_{ij}) exists so that $Q(\xi)$ takes either one of the forms

$$Q(\eta) = \eta_1^2, \qquad Q(\eta) = \eta_1^2 + \eta_2^2, \qquad Q(\eta) = -\eta_1^2 + \eta_2^2;$$

only in the second case can (6) be called parabolic. Furthermore such regions might have finite volume in three-space, rather than reducing to a two-dimensional surface which separates a region of ellipticity from one of hyperbolicity.

Using a different terminology, system (6) is degenerate iff the matrix A has at least one zero eigenvalue, it is elliptic iff all the eigenvalues of A are strictly of one sign (strictly positive, say) and it is hyperbolic iff the eigenvalues of A are non-zero, but have different signs (one negative and two positive, say). The eigenvalues λ of A are the roots of its characteristic polynomial,

$$P(\lambda) = \det(A-\lambda I),$$

where I is the 3x3 identity matrix. Hence system (6) will be non-degenerate iff matrix A is non-singular, i.e., iff the zeroth order term of $P(\lambda)$ is non-zero,

$$\det A \neq 0.$$

System (6) will be elliptic iff A_{11} and $A_{11}A_{22} - A_{12}^2$ have the same sign as det A (Hildebrand, 1965, pp. 50 - 52). Thus the ellipticity conditions for (6) are

$$(9a) \qquad A_{11} \det A > 0,$$

$$(9b) \qquad (A_{11}A_{22} - A_{12}^2) \det A > 0.$$

We have not succeeded in expressing the quantities A_{11}, $A_{11}A_{22} - A_{12}^2$ and det A in terms of the right-hand sides of (6a),(6b) or in any other way that does not depend on the derivatives of u,v. However, given a finite-difference approximation to these derivatives, it is a straightforward, albeit lengthy, computation to evaluate these quantities from the expressions for (A_{ij}) obtained by expanding (8).

We see that the situation is more complicated for system (3) than it was for (3.1). Still system (6) represents a definite possibility for initialization by compatible balancing in a primitive-equation model. Numerous problems in the implementation of the proposed procedure remain open: we shall briefly discuss some of them in the final section.

5. Concluding Remarks

We have shown that the horizontal wind field is synoptically determined by the mass field and its first two time derivatives, at least in a number of mathematical models describing the behavior of the atmosphere. The diagnostic equations for the velocity field were in each case derived from the prognostic equations of the model by purely mathematical manipulations; we did not make any additional physical assumptions which would involve the omission of certain terms in the equations. Therefore, initial states with prescribed mass field, and with the wind field given by the corresponding diagnostic equations, should generate time-dependent solutions of the model under consideration in which the features representing meteorologically significant phenomena predominate.

From the theoretical viewpoint, our results seem to explain the partial success of updating as well as the difficulties encountered. According to the analysis presented here, winds are determined by the mass field and by some information contained in its past history; on the other hand, the relevant information is not exploited in an optimal way by the updating techniques. We feel that the systematic procedure for initialization by compatible balancing as outlined here is an interesting alternative both to non-synoptic initialization and to more traditional synoptic techniques.

Two kinds of difficulties are to be expected in the practical implementation of this approach to initialization: (a) observational, and (b) numerical.

(a) The observational difficulties have mainly to do with the fact that synoptic coverage of the mass field, though considerably better than that for the wind field, is still not complete. Satellite coverage has proved helpful, but some problems in the processing of satellite data have proved to be more serious than expected (Bengtsson, 1975). Part of these problems are also of a mathematical nature, having to do with the ill-posedness of a certain integral equation; it is hoped that known techniques for dealing with such problems could lead to the computation of better vertical temperature distributions from satellite data.

(b) The numerical difficulties to be expected in this initialization approach have to do basically with the discrepancy between the solutions of numerical models and those of the mathematical models they approximate. The wind field given by the solution of the proposed diagnostic differential equations is compatible with the prescribed mass field with respect to the differential equations of the prognostic model; but in practice one can solve only a set of difference equations which approximate the diagnostic equations we derived. The problem then arises of obtaining diagnostic difference equations directly from the prognostic difference equations of a given numerical model. This task is facilitated by the fact that, in each model we discussed, one of the equations of the diagnostic system was actually identical to one of the equations of the prognostic system; taking the same discretized version of this diagnostic equation for determining the wind field as is used in the time integration of the prognostic numerical model will solve at least part of the problem. The second discrete equation may then also be obtained more readily in a compatible way.

This approach to the numerical difficulties also suggests a solution to the problem mentioned in Section 2 of calculating the required time derivatives of the mass field from discrete (as opposed to continuous) observations. Since eventually the initialization will be done for a discrete model, it is natural to take time differences instead of time derivatives. It is true that the time intervals used in present explicit numerical models are of the order of minutes, while synoptic maps are available at time intervals of the order of hours (Bengtsson, 1975). On the other hand, we are interested in mass-field time-difference information which is filtered so as to make the meteorologically significant phenomena dominant in the initial state obtained with the help of this information. Therefore it is desirable for our purposes to take time differences based on mass-field maps processed by standard synoptic procedures from raw data, and available at time intervals characteristic of changes in the meteorologically significant large-scale motions of the atmosphere, rather than at time intervals characteristic of changes in inertia-gravity wave motions. There is certainly a need for investigations on optimum time intervals and differencing methods for the time derivatives required by the initialization method presented here.

The major problem, however, in implementing this method is that of constructing solutions of a nonlinear diagnostic system of mixed type, and of devising numerical schemes for doing so which will be sufficiently efficient in order to make the method practically useful. We hope to be able soon to report on progress in this direction.

REFERENCES

1. Bengtsson, L., 1975. 4-Dimensional Assimilation of Meteorological Observations, GARP Publications Series, No. 15, WMO-ICSU, Geneva, 76 pp.

2. Bers, L., 1958. Mathematical Aspects of Subsonic and Transonic Gas Dynamics, John Wiley & Sons, New York, 164 pp.

3. Bleck, R., 1974. Short-range prediction in isentropic coordinates with filtered and unfiltered numerical models, Mon. Wea. Rev., 102, 813-829.

4. Charney, J. G., 1955. The use of the primitive equations of motion in numerical prediction, Tellus, 7, 22-26.

5. ————————, R. G. Fleagle, V. E. Lally, H. Riehl and D. Q. Wark, 1966. The Feasibility of a Global Observation and Analysis Experiment, Publication 1290, Natl. Acad. Sci./Natl. Res. Council, Washington, D. C., 172 pp.

6. ————————, M. Halem and R. Jastrow, 1969. Use of incomplete historical data to infer the present state of the atmosphere, J. Atmos. Sci., 26, 1160-1163.

7. Cole, J. D., 1968. Perturbation Methods in Applied Mathematics, Blaisdell, Waltham, Mass., 260 pp.

8. Courant, R., and D. Hilbert, 1962. Methods of Mathematical Physics, vol. II, Wiley-Interscience, New York, 830 pp.

9. Davies, H. C., and R. E. Turner, 1975. Dynamical relaxation: A technique for updating prediction models, to appear in the Monthly Weather Review.

10. Eliassen, A., and E. Kleinschmidt, 1957. Dynamic Meteorology, in the Encyclopedia of Physics, vol. XLVIII, edited by J. Bartels, Springer, Berlin, pp. 1-154.

11. Ellsaesser, H. W., 1968. Comparative test of wind laws for numerical weather prediction, Mon. Wea. Rev., 96, 277-285.

12. Elvius, T., and Sundström, 1973. Computationally efficient schemes and boundary conditions for a fine-mesh barotropic model based on the shallow-water equations, Tellus, 25, 132-156.

13. Fjørtoft, R., 1962. A numerical method of solving certain partial differential equations of second order, Geofys. Publik., 24, 229-239.

14. Garabedian, P. R., 1964. Partial Differential Equations, John Wiley & Sons, New York, 672 pp.

15. Ghil, M., 1973. On Balance and Initialization, Report IMM-400, Courant Institute of Mathematical Sciences, New York University, 32 pp.

16. Haltiner, G. J., 1971. Numerical Weather Prediction, John Wiley & Sons, New York, 317 pp.

17. Hildebrand, F. B., 1965. Methods of Applied Mathematics, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 362 pp.

18. Hinkelmann, K., 1951. Der Mechanismus des meteorologischen Lärmes, Tellus, 3, 285-296.

19. ——————, 1962. Noncharacteristic filtering of meteorological noise waves, Beitr. Phys. Atmos., 35, 252-276.

20. Hollmann, G., 1966. Zur Frage neuer diagnostischer Beziehungen zwischen Wind - and Druckfeld (Balancegleichungen) in einer barotropen Atmosphäre mit divergenter Strömung, Beitr. Phys. Atmos., 39, 99-111.

21. Houghton, D. D., 1968. Deviation of the elliptic condition for the balance equation in spherical coordinates, J. Atmos. Sci., 25, 927-928.

22. —————— and W. Washington, 1969. On global initialization of the primitive equations: Part I, J. Appl. Meteor., 8, 726-737.

23. Jastrow, R., and M. Halem, 1973. Simulation studies and the design of the first GARP global experiment, Bull. Amer. Meteor. Soc., 54, 13-21.

24. Jameson, A., 1975. Numerical solution of nonlinear partial differential equations of mixed type, to appear in Numerical Solution of Partial Differential Equations - III, edited by I. Babuška, Academic Press, New York.

25. Kasahara, A., 1972. Simulation experiments for meteorological observing systems for GARP, Bull. Amer. Meteor. Soc., 53,252-264.

26. Kistler, R. E., and R. D. McPherson, 1975. On the use of a local wind correction technique in four-dimersional data assimilation, Mon. Wea. Rev., 103, 445-449.

27. Kreiss, H.-O., and J. Oliger, 1973. Methods for the Approximate Solution of Time Dependent Problems, GARP Publication Series, No. 10, WMO-ICSU, Geneva, 107 pp.

28. Krishnamurti, T. N., 1968. A diagnostic balance model for studies of weather systems of low and high latitudes, Rossby number less than 1, Mon. Wea. Rev., 96, 197-207.

29. Mesinger, F., 1972. Computation of the wind by forced adjustment, J. Appl. Meteor., 11, 60-71.

30. Miranda, C., 1970. Partial Differential Equations of Elliptic Type, 2nd ed., Springer, New York, 370 pp.

31. Miyakoda, K., and Moyer, R. W., 1968. A method of initialization for dynamical weather forecasting, Tellus, 20, 115-128.

32. Morel, P., G. Lefevre and G. Rabreau, 1971. On initialization and non-synoptic data assimilation, Tellus, 28, 197-206.

33. Murman, E. M., and J. D. Cole, 1971. Calculation of plane steady transonic flows, AIAA Journal, 9, 114-121.

34. Nitta, T., and J. B. Hovermale, 1969. A technique of objective analysis and initialization for the primitive forecast equations, Mon. Wea. Rev., 97, 652-658.

35. Paegle, J., and J. N. Paegle, 1974. An efficient and accurate approximation to the balance wind with application to non-elliptic data, Mon. Wea. Rev., 102, 838-846.

36. Pogorelov, A. V., 1964. Monge-Ampère Equations of Elliptic Type, Noordhoff, Groningen, 116 pp.

37. Rellich, F., 1932. Zur ersten Randwertaufgabe bei Monge-Ampèreschen Differentialgleichungen vom elliptischen Typus; differentialgeometrische Anwendungen, Math. Ann., 107, 505-513.

38. Shuman, F. G., 1957. Numerical methods in weather prediction: I. The balance equation, Mon. Wea. Rev., 85, 329-332.

39. Smagorinsky, J., and K. Miyakoda, 1969. The relative importance of variables in initial conditions for numerical predictions, in Proc. WMO/IUGG Symp. Numerical Weather Prediction, Tokyo, 1968, Meteorological Society of Japan, pp. V.1-V.2.

40. Stone, P. H., L.-C. Tsang and D. Schneider, 1973. Balanced winds for assimilation of temperature and pressure data, GISS Res. Rev., NASA, 1973, pp. 160-163.

41. Williamson, D. L., and R. E. Dickinson, 1972. Periodic updating of meteorological variables, J. Atmos. Sci., 29, 190-193.

42. ———————, and A. Kasahara, 1971. Adaptation of meteorological variables forced by updating, J. Atmos. Sci., 28, 1313-1324.